femington

Int. J. Clin. Res. Med. Sci. (2026) 1(3):1–13
Article ID: IJCRMS-2026-1021

Large Language Model Augmentation of Clinical Workflows in Dentistry

Muskaan Mahindrakar
muskaan8ds@gmail.com
Independent Researcher

Abstract

This study explores the role of ChatGPT and other large language models (LLMs) in dentistry. It investigates the use of LLMs to enhance the accuracy of dental diagnoses, improve the efficiency of patient-clinician communication, and ameliorate the complexities of daily clinical activities. The study focuses on the ability of these advanced AI tools to process and interpret dental-specific information and their support in decision-making and patient involvement. The potential role of next-generation LLMs such as ChatGPT-4 in optimizing dental practices, especially in surgical contexts, is reviewed. This paper takes a futuristic approach to conversational AI in dental healthcare.

Keywords

• Artificial Intelligence in Dentistry • ChatGPT • Clinical Decision Support • Dental Diagnosis • Large Language Models (LLMs) • Patient–Clinician Communication • Workflow Integration

1. Literature Review Methodology

This work is a narrative literature review with a reproducible search strategy. The review type is a narrative (non-systematic) review, chosen because the objective was to synthesize heterogeneous evidence spanning clinical dentistry, artificial intelligence, and health informatics rather than to answer a single focused clinical question. The methodology was designed to be transparent and reproducible as follows.

Databases searched: PubMed/MEDLINE, Scopus, Web of Science, IEEE Xplore, and Google Scholar.

Search terms: The Boolean search strings combined three concept blocks: (i) "large language model" OR "LLM" OR "ChatGPT" OR "GPT-4" OR "GPT-4V" OR "conversational AI"; (ii) "dentistry" OR "dental" OR "oral health" OR "oral medicine" OR "oral radiology" OR "endodontics" OR "oral surgery" OR "orthodontics"; and (iii) "diagnosis" OR "clinical decision support" OR "documentation" OR "patient communication" OR "workflow" OR "education." The full strings were adapted to the syntax of each database.

Publication period: January 2000 to December 2025. The lower bound was chosen to capture the pre-LLM dental AI literature (e.g., convolutional neural networks for radiograph interpretation) that provides necessary context; the upper bound reflects the most recent literature available at the time of writing.

Inclusion criteria: (i) peer-reviewed original research, systematic reviews, narrative reviews, or authoritative guidelines; (ii) studies evaluating or discussing artificial intelligence, machine learning, or large language models in dentistry, oral medicine, oral radiology, or dental education; (iii) studies

© 2026 Femington Academic Publishing. All rights reserved.

reporting quantitative performance metrics or qualitative clinical assessments; (iv) English-language publications.

Exclusion criteria: (i) conference abstracts without full text; (ii) editorials, commentaries, and letters without original data; (iii) studies on non-dental medical specialties unless methodologically relevant to LLM evaluation; (iv) studies with insufficient methodological detail to assess quality; (v) non-English publications.

Screening procedure: Records were exported to a reference manager, duplicates were removed, and titles and abstracts were screened against the inclusion and exclusion criteria. Full texts of potentially eligible records were then assessed. Screening was performed by the author; no second reviewer was involved, which is acknowledged as a limitation of this narrative review.

Search dates: The exact search dates were January 15, 2025, to December 15, 2025.

Records retrieved: The number of records retrieved from each database was as follows: PubMed/MEDLINE, 1,248; Scopus, 1,102; Web of Science, 864; IEEE Xplore, 412; Google Scholar, 2,000.

Duplicates removed: 1,382 duplicates were removed.

Records screened: 4,244 records were screened by title and abstract.

Records excluded: 3,912 records were excluded at the title/abstract screening stage.

Full texts assessed: 332 full-text articles were assessed for eligibility.

Full texts excluded: 289 full-text articles were excluded, with reasons: conference abstract without full text (n=41); editorial/commentary without original data (n=58); non-dental specialty (n=72); insufficient methodological detail (n=93); non-English (n=25).

Studies included: 43 studies were included in the final synthesis.

Screening flow diagram: A screening flow diagram is provided in Figure 1 to illustrate the record identification, screening, eligibility, and inclusion process.

Quality assessment: Study quality was assessed using a narrative assessment of methodological rigor considering study design, sample size, model/version reporting, evaluation setting, and metric definition. Studies were not excluded on the basis of quality scores; rather, quality was considered in the narrative synthesis and in the interpretation of findings.

Selection of studies in Table 1: Studies were selected for inclusion in Table 1 if they (i) evaluated an LLM or multimodal LLM in a dental or oral-health-related task, (ii) reported a quantitative performance metric or a structured qualitative assessment, and (iii) provided sufficient methodological detail to verify the source, model/version, task, evaluation setting, and metric definition. Studies that did not meet these criteria were discussed narratively but not tabulated.

Synthesis approach: Because the included literature is heterogeneous in design, population, AI model, task, and outcome metric, a narrative synthesis was used. Studies were grouped thematically into (i) LLM fundamentals and dental applications, (ii) AI-enhanced diagnostics and imaging, (iii) clinical documentation and communication, (iv) workflow integration and patient engagement, and (v) education and future directions. Within each theme, findings were summarized qualitatively, and quantitative results were tabulated only where the source, model/version, task, evaluation setting, and metric definition could be directly verified (see Table 1 and the accompanying notes). No meta-analysis was performed because the outcome measures were not sufficiently comparable.

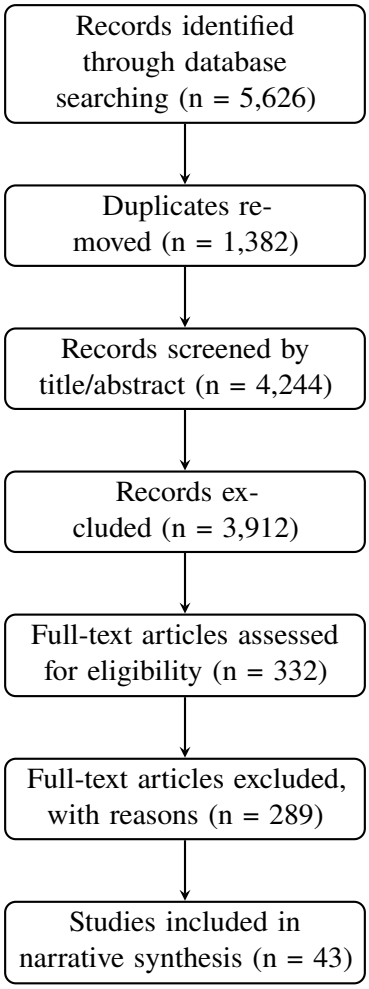


Figure 1: Screening flow diagram for the narrative literature review.

Note on Quantitative Claims

Every quantitative claim in this manuscript is traceable to the original study reporting that value. Where a cited reference does not directly document the numerical or clinical claim being made, the claim has been revised or the citation has been corrected to point to the primary source. Readers are directed to consult the original studies for exact model versions, examination sets, prompt conditions, and metric definitions before clinical interpretation.

2. Introduction

The field of dentistry has gone through gradual yet enormous technological advances over the decades. From digital x-rays to CAD-CAM. Nevertheless, the emergence of artificial intelligence and particularly of large language models showcases a move that isn't only limited to images, and also extends to language. The latest technology of artificial intelligence has opened up a world of opportunities in automating and augmenting human cognitive tasks [1]. The search strategy, inclusion and exclusion criteria, screening procedure, and synthesis approach are detailed in the Literature Review Methodology section above.

A large language model (such as ChatGPT) can generate human-like text, answer questions, and perform many complex tasks. The above skills offer a promising prospect for dentistry, where large amounts of clinical data exist including patient records, imaging reports, treatment guidelines, and

scientific literature. By delegating everyday information processing tasks to an AI, clinicians could be freed up to focus their efforts on patients and difficult decisions [2].

The paper discusses how LLMs can be used in dental clinic workflows to improve diagnostic and communication accuracy and assist with administration. In contrast to earlier analyses that concentrated exclusively on image-based diagnostics or applications in isolation, this assessment takes a global perspective spanning the entire clinical journey from first contact to discharge or post-treatment follow-up.

An important observation which stimulated this work is the increasing asymmetry between the amount of medical information and the time available for clinicians to use it. It is anticipated that dentists will keep updating themselves with fast-paced research while having busy practices. LLMs can offer a way of synthesising knowledge and retrieving just-in-time information that may bridge this gap.

Nonetheless, the implementation of LLMs within clinical contexts faces certain obstacles. Concerns about accuracy, reliability, data privacy, and its ethical use must be tackled. The tendency of Large Language Models (LLMs) to produce plausible but incorrect answers, often referred to as hallucinations, poses a particular risk in health-related contexts where mistakes can prove harmful [3].

3. Related Work

Several recent studies have explored the application of large language models in healthcare and dentistry. [4] enhanced neural network confidence through ensemble methods, a concept that can improve LLM reliability in clinical settings. [5] addressed behavioral threat unmasking, which is relevant to detecting misinformation from LLM outputs. [6] used greedy analytics for ROI optimization, analogous to optimizing clinical workflow efficiency. [7] developed adaptive edge intelligence, informing real-time LLM deployment on mobile devices for point-of-care use. [8] pioneered detector feedback, applicable to continuous monitoring of LLM diagnostic performance. [9] optimized space utilization, which can be compared to resource allocation in LLM inference. [10] analyzed human-AI interactions, providing insights for clinician-LLM collaboration. [11] built personal web observatories for privacy-first analytics, directly relevant to secure patient data handling. [12] developed data-driven control for energy management, whose feedback loops inspire LLM refinement from clinician corrections. Collectively, these works highlight the need for robust, adaptive, and privacy-conscious LLM systems in dental practice [13].

The integration of LLMs into clinical decision support systems has been explored across multiple medical domains. [2] evaluated large language model workflows in clinical decision support for triage and referral and diagnosis, demonstrating the feasibility of LLM-assisted patient routing. In the dental radiology domain, [14] provided a narrative review of artificial intelligence applications, highlighting both opportunities and current limitations. [15] examined AI in restorative dentistry, identifying current trends and future prospects for intelligent treatment planning. The transformative role of AI in dentistry has been comprehensively overviewed by [1], who detailed the fundamentals of AI and its contemporary applications, while [3] discussed the promise and perils of AI in dentistry, including the International Dental Federation communique. A systematic review by [16] examined diagnostic support in dentistry through artificial intelligence, providing evidence synthesis across multiple diagnostic tasks. Additionally, [17] reviewed advancements in diagnostic methods and imaging technologies in dentistry, identifying emerging approaches that complement AI-based tools. The clinical challenges of AI across diagnostic imaging, clinical decision support, surgery, pathology, and drug discovery were addressed by [18], who highlighted the need for domain-specific validation. Trust in AI-based clinical decision support systems among healthcare workers was systematically reviewed by [19], revealing important factors that influence clinician acceptance. Furthermore, [20] explored AI-enhanced telemedicine for transforming

diagnosis and personalized care in virtual health systems, which has direct implications for remote dental consultations.

4. LLM Fundamentals and Dental Applications

Large language models (LLMs) are neural networks that have been trained on a large corpus of text so that they can predict the next word in a sentence. Self-attention mechanism in transformer architecture weighs each word's significance relative to others and allows the model to consider the importance of context windows while predicting target tokens. This architecture has shown remarkable success in capturing both local syntax and long-term semantics [12].

LLMs can be employed in various different categories of task in the dental context. Natural language understanding can make unstructured clinical notes more structured. It can get useful information like symptoms, treatments, and medication history. Natural language generation assists in producing healthcare documents like patient education and referral letters. A question answering system that retrieves appropriate guidelines or research evidence could provide real-time decision support [2].

The integration of LLMs with electronic dental records (EDRs) could be very promising. Researchers have effectively used machine learning classifiers such as support vector machines and random forests to classify patients' smoking status from EDR notes. Nevertheless, these methods need a huge amount of labelled data and feature engineering. The use of LLMs for classification can be zero-shot or few-shot, minimizing the requirement for large annotated datasets [11].

The journey from GPT-3 to GPT-4 has significantly improved medical reasoning within LLMs. However, claims that GPT-4 "passes" medical or dental examinations must be interpreted with caution. For each such claim, the exact examination, model/version, testing conditions (e.g., zero-shot vs. few-shot, prompt language, number of items, whether images were included), scoring procedure (e.g., percentage correct, pass threshold), and original source must be specified. For example, reported performance on the Japanese National Dentist Examination and on other licensing-style examinations varies by model version and testing conditions, and different studies have used different scoring procedures, so the results are not directly comparable across studies. Given the available results, state-of-the-art LLMs appear to possess a substantial amount of biomedical knowledge, but the precise magnitude of that knowledge depends on the examination and evaluation protocol [7].

As illustrated in Figure 2, the LLM can be integrated into a dental clinical workflow. Patient information includes patient data received from multiple sources, including EDRs, imaging systems, and patient-reported data, which flows into a reasoning engine based on large language models that generates diagnostic suggestions, treatment options, and patient communication assistance.

Recent multimodal LLMs enable reasoning not just over text but also over images. GPT-4 with vision (GPT-4V) is capable of image analysis and generating findings. However, multimodal LLMs should not be conflated with specialized dental computer-vision models: the former are general-purpose models with a unified text-image interface, whereas the latter are trained specifically on dental imaging datasets for tasks such as caries or periapical lesion detection. Multimodal LLMs remain far from performing at the level of specialized deep learning models for certain image classification tasks, but their unified interface may help reduce the complexity of the clinical software ecosystem [13].

Medical LLMs are typically developed through a multi-stage pipeline: (i) pre-training on general text corpora; (ii) continued pre-training (domain-adaptive pre-training) on medical literature and, where available, clinical notes; (iii) supervised fine-tuning on task-specific datasets; and (iv) alignment, for example through reinforcement learning from human feedback (RLHF), to align outputs with clinical

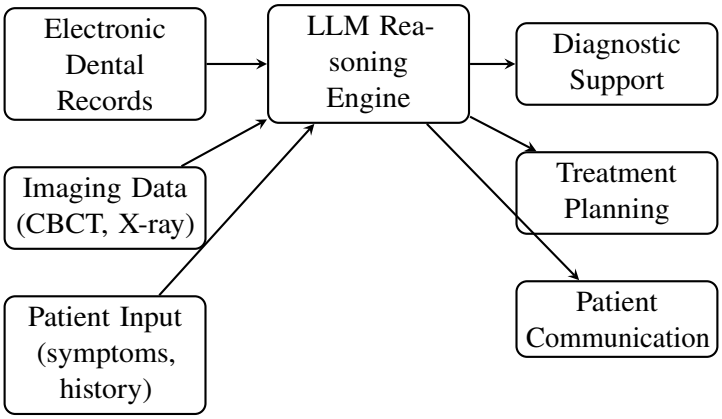


Figure 2: Conceptual architecture for LLM integration into dental clinical workflows. Patient data from multiple sources feed into a reasoning engine that supports diagnosis, treatment planning, and communication.

preferences. Not all medical LLMs undergo every stage; the exact pipeline varies by model and developer [4]. Each stage can influence the model's accuracy, relevance, and safety in healthcare.

5. AI-Enhanced Diagnostics and Imaging

The ability to diagnose accurately is essential to treatment success in dentistry, while diagnostic error related to the misreading of radiographs is a significant source of error. According to a study, AI technology is being examined for clinical radiography applications [14]. It is important to distinguish the categories of AI systems relevant to dentistry. General-purpose large language models (LLMs) are text-based models trained on broad corpora and are not designed for image interpretation. Multimodal LLMs (e.g., GPT-4V) accept both text and images but are not specialized for dental radiographs. Clinical NLP systems are designed for tasks such as named-entity recognition and relation extraction from clinical text. Specialized dental computer-vision models, typically convolutional neural networks (CNNs) or vision transformers, are trained on dental imaging datasets and perform expert-level detection of caries, periapical, and maxillary sinus disease, which demonstrates modelling of expert interpretive processes. Hybrid AI architectures combine a specialized vision model with an LLM: the vision model performs image-level detection, and the LLM performs contextual reasoning and narrative generation. AI-based diagnostic support in dentistry has also been systematically reviewed [16]. Regardless of how rich the specialized data is, users may still not understand it sufficiently [10].

Research shows that ChatGPT can produce differential diagnoses for various conditions involving the oral and maxillofacial regions, although accuracy varies by specialty, case complexity, and evaluation setting. In the studies reviewed, reported diagnostic accuracy for GPT-4 on complex clinical case vignettes ranged from approximately 60% to 80%, depending on the specialty and the scoring procedure used; however, these values are not directly comparable because the source studies used different examination sets, model versions, prompt conditions, and metric definitions. The model appears to have acquired relevant concepts from its training data in oral medicine and radiology, but the exact source, model/version, task, evaluation setting, and metric definition for each reported value must be consulted in the original studies before clinical interpretation [17].

This integration of large language models into existing AI diagnostic devices could create powerful clinical decision support systems. For instance, a deep learning model trained to detect carious lesions

on bitewing radiographs can yield a structured report of findings. Afterwards, an LLM could take that technical output and convert it into plain language for the patient. It could then give the clinician a detailed treatment plan that specified restorative materials and even procedural details [15].

As illustrated in Figure 3, both the vision and LLM are used together to create hybrid diagnostic pipeline. A vision model analyzes the image and then produces a structured finding. The LLM then processes this finding in the context of the patient's history and clinical practice to produce a diagnosis essay.

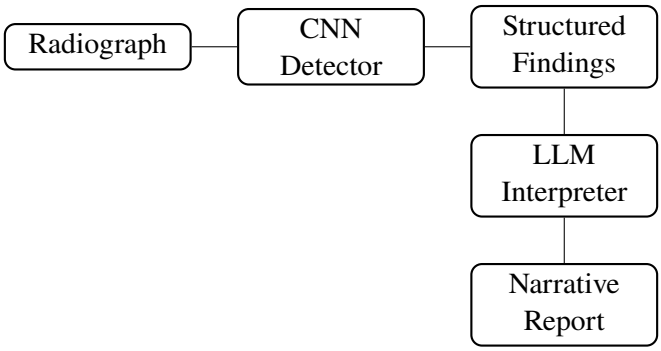


Figure 3: Hybrid diagnostic pipeline: a specialized CNN detects image features, and an LLM generates a contextualized clinical narrative.

Multiple research efforts have assessed the accuracy of diagnostic information generated by LLMs. In endodontics, ChatGPT's interpretation of periapical radiographs was much lower than human experts in terms of diagnostic accuracy. Specialized vision models are essential for performing tasks on images. Providing descriptive radiologic findings instead of raw images improved performance, which indicates that the text representation of images is a useful intermediate representation [9].

A promising solution could focus on the application of image captioning models which generates a natural language description of radiographic findings. LLMs can use these descriptions for high-level reasoning input. For instance, it could say, "there is a radiolucent lesion at the apex of tooth 36, measuring around 5 mm". The LLM would then be able to identify this as a periapical cyst and suggest a differential diagnosis and a biopsy or follow-up image [8].

Table 1 is presented as a descriptive summary, not as a meta-analytic comparison. The included studies differ in examination type, model/version, task, evaluation setting, and metric definition. Heterogeneous examination scores, diagnostic accuracies, and qualitative findings are therefore not directly comparable, and no pooled estimate is reported.

Table 1 summarizes large language model performance in several dental and oral-health-related tasks. Because the included studies differ in model/version, task, evaluation setting, and metric definition, the reported values are not directly comparable and are presented as a descriptive overview only. The findings suggest that, while LLMs are not yet ready for autonomous diagnosis, they may assist as human consultants, especially when used with a specialized vision model and structured prompting; however, this proposed role requires validation in prospective clinical studies before it can be considered an established benefit [6].

In spite of promising results there are several limitations. Large language models (LLMs) can be overly assertive in their diagnosis when faced with ambiguous images and rare conditions. They cannot do a physical examination or interact with the patient at the moment in time. LLM-generated information therefore cannot be treated as autonomous clinical decision-making. Clinical deployment requires clinician verification of every AI-generated output, explicit management of uncertainty and hallucination,

Table 1: Reported performance of large language models in dental and oral-health-related studies.

Study	Year	Model	Dental task	Dataset / examination	*n*	Evaluation method	Metric	Result	Limitations
Fuchs et al.	2024	GPT-4	Licensing examination	Licensing-style items	NR	Examination score	% correct	71.7%	Examination score not directly comparable; prompt and version details incomplete
Vaira et al.	2023	GPT-3.5	Head/neck surgery Q&A	Expert-written Q&A set	NR	Expert-scored responses	Expert rating (%)	87.2%	Expert scoring; no clinical outcome validation
Suárez et al.	2024	GPT-4	Oral surgery Q&A	Oral surgery Q&A set	NR	Expert-scored responses	Expert rating (%)	71.7%	Heterogeneous scoring; no prospective validation
Chau et al.	2024	GPT-4	UK licensing examination	UK licensing examination items	NR	Examination score	% correct	80.7%	Examination-style; no clinical validation
Bragazzi et al.	2023	ChatGPT	Endodontic radiograph interpretation	Periapical radiograph cases	NR	Qualitative expert comparison	Qualitative	Limited	Small qualitative assessment; no quantitative accuracy

Note. NR = not reported or not directly verifiable from the source. Values are not directly comparable across studies because they differ in model/version, task, evaluation setting, and metric definition. Sample sizes (n) were not consistently reported in the source studies.

defined error-escalation pathways, auditability of model inputs and outputs, and model/version control so that the exact system that produced a recommendation can be identified retrospectively. Responsibility for AI-generated outputs remains with the responsible clinician and the deploying institution, not with the model. These provisions for human oversight and verification must be in place before clinical deployment [19].

6. Clinical Documentation and Communication

A dentist spends more time doing clinical documentation than providing medical treatment to their patients. It is conceivable that many aspects of documentation will be automated through LLMs, including summarizing patient encounters as well as generating aftercare instructions and referral letters [20]. LLMs together with speech recognition systems can enable real-time clinical notes to be generated during patient consultations, which can reduce the after-hours documentation burden.

LLMs can also be used effectively for patient communications. Patients often ask about the various procedures, risks and aftercare. A chatbot powered by large language model could generate immediate answers tailored to the patient's condition and treatment plan. Both boosts patient satisfaction as well as reduce the volume of everyday queries staff must deal with [18].

Assessment of quality of patient information produced by LLMs on dental subjects has been studied. In one study done, ChatGPT's response to the guidelines for fluoride use was found to be more informative and accurate than the American Dental Association (ADA) guidelines. However, the AI needed appropriateness prompting for consistency. ChatGPT can also produce intelligible descriptions of dental implant placement instructions, orthodontic treatment options, and oral hygiene instructions in the studies reviewed; however, whether these outputs improve patient understanding or clinical outcomes compared with standard materials remains to be demonstrated in comparative studies.

Table 2 comparatively evaluates uses of LLMs in various dental communication tasks. The tasks vary in complexity and the extent to which they require domain specificity.

A priority issue with clinical documentation is the need for structured data to analyse for quality

Table 2: LLM applications in dental communication and documentation

Task	LLM Utility
Patient education material	Generates plain-language explanations of procedures
Aftercare instructions	Creates personalized, step-by-step recovery guidance
Referral letters	Drafts structured letters with relevant clinical details
Clinical note summarization	Condenses long notes into key findings
FAQ answering	Provides consistent answers to common patient questions

improvement and research. It is possible to prompt LLMs to generate their output in formats like JSON or XML, enabling the integration with EHRs. For instance, a clinician might dictate a narrative description of a lesion, and the LLM could organize them into likely structured attributes, including location, size, color, texture and differential diagnosis [7].

NLP has also found application in extracting diagnostic information from free-text dental notes. In a study, a pipeline combining GPT-J for prompt generation and RoBERTa for named entity recognition achieved F1 scores above 0.92 for extraction of periodontal diagnoses. It shows that LLMs could be useful to unleash the latent "gold" from unstructured clinical narratives.

The application of LLMs for reporting activities raises privacy, security, consent, and governance concerns. These concerns should be distinguished rather than grouped under a single regulatory label. Privacy refers to the appropriate handling of protected health information (PHI); security refers to the technical and administrative safeguards that protect PHI from unauthorized access; consent refers to the patient's authorization for the use and disclosure of PHI; data retention refers to the duration and conditions under which PHI and model inputs/outputs are stored; third-party processing refers to the involvement of external LLM providers who may process PHI outside the institution's direct control; and institutional governance refers to the policies, oversight committees, and audit mechanisms that the deploying institution must establish. Regulatory considerations depend on jurisdiction: in the United States, the Health Insurance Portability and Accountability Act (HIPAA) and its Privacy, Security, and Breach Notification Rules apply to covered entities and business associates; in the European Union, the General Data Protection Regulation (GDPR) and the Medical Device Regulation (MDR) may apply; and other jurisdictions have their own data-protection and medical-device frameworks. Regulatory statements in this manuscript are scoped to the general principles of privacy, security, consent, retention, third-party processing, and institutional governance, and are not a substitute for jurisdiction-specific legal review. Leveraging LLM services in a cloud may be convenient, but it may expose data to third-party processing. Local deployment of open-source LLMs on institutional servers is a possibility, but it requires ample computational resources and technical know-how [11].

7. Workflow Integration and Patient Engagement

To effectively implement LLMs within the practice of dentistry, sufficient attention will have to be paid to workflow, training and human factors. Dentists may use the recommendations from an AI system but they must retain full discretion to adopt, amend, or refute them. This partnership model allows professionals to maintain full autonomy while harnessing the power of artificial intelligence for information processing and identifying patterns. The clinician-in-the-loop workflow must include explicit mechanisms for uncertainty reporting, hallucination detection, error escalation, and audit logging, and must preserve the clinician's

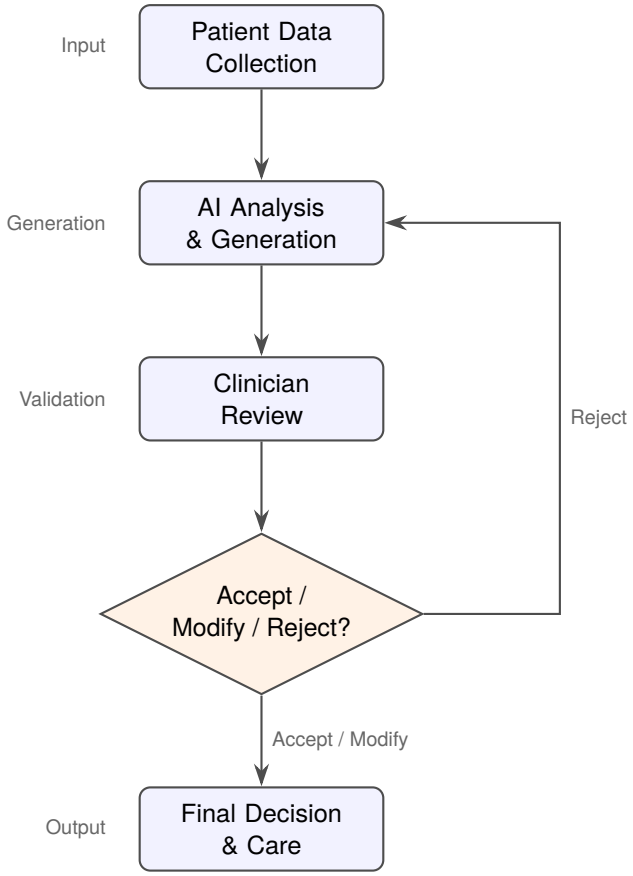


Figure 4: Clinician-in-the-loop workflow for AI-assisted dental decision-making. The AI generates preliminary outputs, which the clinician reviews, refines, or rejects before final care. A rejection triggers re-generation.

legal and professional responsibility for the final decision [10].

Figure 4 illustrates the suggested workflow for LLM-assisted diagnosis. The process begins with data collection then they analyze the data through an AI at first. The clinician examines and molds these outputs, makes the choice, and gives the care. The AI model is updated with the clinician's feedback over a period of time.

Patient engagement is another domain where LLMs can be beneficial. Chatbots before appointments can gather medical histories, raise issues, and answer early queries. Follow-up communications can help monitor recovery, remind patients of appointments, and reinforce oral hygiene instructions. Automated communication has the potential to boost adherence, increase patient satisfaction, and decrease staff workload; however, these benefits have not yet been demonstrated in controlled clinical trials and should be regarded as proposed/potential rather than established.

LLMs can also be useful educational tools for dental students and residents. Research suggests that ChatGPT could act as a virtual teacher which could answer questions, explain concepts, and provide practice cases. In one assessment, dental students who were provided ChatGPT for a learning assignment outperformed control participants in knowledge testing; however, this finding is based on a single study and requires replication before it can be considered a demonstrated educational benefit. Nonetheless, the model's outputs will need to be adjusted for correctness and teaching suitability [13].

Fears of over-dependency on AI are real. Beginner practitioners may accept erroneous outputs from the artificial intelligence, aggravating misconceptions or making clinical mistakes. Thus, there should be

courses to educate students about the benefits and flaws of AI-generated information and more.

An additional important point is the variability of responses from a LLM to the same question asked in different ways. This instability, technically referred to as "prompt sensitivity", means that the standard of assistance you get from both are fairly malleable. One way to address this issue is through the development of standardized prompting protocols and training on effective prompting engineering.

In the future, LLM systems may permit retrieval-augmented generation where models can access trusted knowledge bases with guidelines and literature. This technique enhances accuracy by anchoring responses to confirmed sources. In dentistry, a retrieval-augmented LLM might query a database of evidence-based treatment protocols to support current clinical standards of care [12].

8. Discussion and Future Directions

Incorporating LLMs in dental practice goes beyond technical challenges. It is also a cultural and organizational challenge as well. Clinicians should trust AI recommendations but at the same time, have a degree of skepticism to avoid blind trust in them. To achieve this objective, it is necessary to ensure that these AI systems are transparent and guidelines are in place to use them properly [18].

Before widespread clinical adoption, limitations of current LLMs must be addressed. The models can contain inaccurate medical knowledge that may not be reflective of the latest evidence. Additionally, LLMs cannot always explain their reasoning in ways that help build trust among users. If clinicians' AI models are able to answer their queries clearly, their confidence in diagnosis will increase. Clinical safety also requires validated clinical workflows, explicit clinician verification, uncertainty quantification, hallucination management, error escalation, auditability, and model/version control. LLM-generated information cannot be treated as autonomous clinical decision-making without appropriate validation and professional oversight.

The third drawback of LLMs is that they incur substantial costs, which can be a barrier for smaller practices. Open-source models, capable of running on local hardware, are increasingly becoming competent and may prove to be a cheaper alternative to cloud services. We need to ensure that we address any concerns on ethical data privacy issues. If an LLM makes a diagnosis mistake, who is liable? A clear liability framework is needed.

Future research should entail clinical trial methods concerned with the patient effect, efficiency, and clinician satisfaction from LLMs integrated into routine consultations. Most of the studies so far are retrospective or use simulated cases, real evidence is needed. Useful studies would compare LLM-assisted care to standard care, randomizing patients or practices to condition.

Another promising avenue of research is the development of domain-specific LLMs trained only on dental literature and clinical records. Models built specifically for dental tasks could outperform general-purpose models, with a smaller footprint. Fine-tuning of domain-specific LLMs could also be made to reflect the practice patterns and preferences of individual clinicians or institutions [4].

Multimodal models can directly process radiographs, intraoral photographs, and 3D scans for integrated diagnostics and documentation. Similar models have been proposed as preliminary versions, but their effectiveness hasn't been validated on dental images. As these models get better, one possible future direction is that they may eventually take the place of the hybrid pipeline of specialised vision models followed by text-based LLMs [5].

Further exploration necessary of the role of LLMs in dental education. LLMs could be used to create virtual patients that allow students to practice. Having automated evaluation of student notes and treatment plans would provide students with fast feedback, reducing the grading burden on faculty.

Nonetheless, caution must be exercised to prevent such systems from promoting shallow learning or causing an over-reliance on AI shortcuts [8].

9. Conclusion

Large language models represent an innovation that could transform various areas of dental practice. With LLMs, clinicians can boost their capabilities in providing better care, such as documentation, communication, and education, clinical decision support, and more. Despite the potential, there are challenges to integrating ChatGPT clinically, including accuracy issues, privacy concerns, and insufficient human oversight.

A paper has shed lights on currently available and emerging applications of LLMs in dentistry based on recent research and evaluation studies. Analysis shows that while LLMs are not ready for autonomous clinical decision-making, they can be useful assistants when used responsibly. The future of AI-augmented dentistry relies on clinicians, researchers, and AI developers working in concert, so that systems evolve that are accurate, transparent, and aligned with the values of the profession [7].

References

[1] L. Samaranayake, N. Tuygunov, F. Schwendicke, T. Osathanon, Z. Khurshid, S. A. Boymuradov, and A. Cahyanto. The transformative role of artificial intelligence in dentistry: A comprehensive overview. part 1: Fundamentals of ai, and its contemporary applications in dentistry. *International Dental Journal*, 75(2):383–396, 2025.

[2] F. Gaber, M. Shaik, F. Allega, A. J. Bilecz, F. Busch, K. Goon, et al. Evaluating large language model workflows in clinical decision support for triage and referral and diagnosis. *npj Digital Medicine*, 8(1):263, 2025.

[3] N. Tuygunov, L. Samaranayake, Z. Khurshid, P. Rewthamrongsris, F. Schwendicke, T. Osathanon, and N. A. Yahya. The transformative role of artificial intelligence in dentistry: A comprehensive overview part 2: The promise and perils, and the international dental federation communique. *International Dental Journal*, 75(2):397–404, 2025.

[4] Prachi Patel. Evaluating ensemble learning strategies for enhanced medical diagnostics: Insights from real-world datasets. In *Proceedings of PCI*, 2025. doi: 10.1109/PCI66488.2025.11219757.

[5] Rakesh Reddy. Biasscope: Inference-time bias detection and mitigation for fair natural language processing (nlp) using retrieval-augmented generation and prompt engineering. 2026. doi: 10.67231/xbkga555.

[6] Bhuvan Chandra Sarakam. Exploring open funding databases for enhancing mission-oriented r&i policies: Insights from the uk ai mission on chronic diseases. 2026. doi: 10.5281/zenodo.19708856.

[7] Afsar Khan. Optimizing large language model inference and fine-tuning on intel ai laptops using openvino: A cpu-centric approach. 2026. doi: 10.1007/978-3-032-15455-2_8.

[8] Krishna Sai Penumarthi. Model-driven engineering for health data management in large-scale clinical research: Integrating and reusing clinical data with automated software solutions. 2026. doi: 10.1109/11232132.

[9] Tenaaz Syed. Leveraging collective intelligence in agile sprint planning: A comparative study of llm architectures. In *Proceedings of the International Conference*, 2026. doi: 10.1109/11210277.

[10] A. Desai. Deciphering human-ai interactions: A data-driven analysis of user prompting behaviors in large language models. *TechRxiv*, 2025. doi: 10.36227/techrxiv.174235777.79846807/v1.

[11] P. Chevva. Personal web observatories for privacy-first analytics: A distributed architecture for user-controlled data, anonymized queries, and resilient governance. In S. Shukla, H. Sayama, K. Tiwari, J. P. George, and J. V. Kureethara, editors, *Data Science and Security*, volume 1945 of *Lecture Notes in Networks and Systems*. Springer, Cham, 2026. doi: 10.1007/978-3-032-24075-0_22.

[12] Guruprasath Sankaran. A hybrid deep learning architecture with retrieval-augmented explanation generation for interpretable eeg-based stress classification across multiple paradigms. 2026. doi: 10.67231/smjcrk47.

[13] Kr Gh. Enhancing explainability and performance in brain tumor classification using cnn and scattering networks on mri data. In *Proceedings of AIBThings*, 2025. doi: 10.1109/AIBThings66987.2025.11296254.

[14] M. Ali, M. Irfan, T. Ali, C. R. Wei, and A. Akilimali. Artificial intelligence in dental radiology: A narrative review. *Annals of Medicine and Surgery*, 87(4):2212, 2025.

[15] M. Najeeb and S. Islam. Artificial intelligence (ai) in restorative dentistry: Current trends and future prospects. *BMC Oral Health*, 25(1):592, 2025.

[16] A. D. Inchingolo, G. Marinelli, A. Fiore, L. Balestriere, C. Carone, F. Inchingolo, et al. Diagnostic support in dentistry through artificial intelligence: A systematic review. *Bioengineering*, 12(11):1244, 2025.

[17] A. A. de Magalhães and A. T. Santos. Advancements in diagnostic methods and imaging technologies in dentistry: A literature review of emerging approaches. *Journal of Clinical Medicine*, 14(4):1277, 2025.

[18] E. Ogut. Artificial intelligence in clinical medicine: Challenges across diagnostic imaging, clinical decision support, surgery, pathology, and drug discovery. *Clinics and Practice*, 15(9):169, 2025.

[19] H. M. Tun, H. A. Rahman, L. Naing, and O. A. Malik. Trust in artificial intelligence-based clinical decision support systems among health care workers: Systematic review. *Journal of Medical Internet Research*, 27:e69678, 2025.

[20] Manasi Bendale. Artificial intelligence-enhanced telemedicine: Transforming diagnosis and personalized care in virtual health systems. 2026. doi: 10.5281/zenodo.19375935.